Assessment of water quality variations under non-rainy and rainy conditions by principal component analysis techniques in Lake Doam watershed, Korea

Bal Dev Bhattrai, Sungjin Kwak and Woomyung Heo*

Department of Environmental Engineering, Kangwon National University, Samcheok 245-711, Korea

Abstract

This study was based on water quality data of the Lake Doam watershed, monitored from 2010 to 2013 at eight different sites with multiple physiochemical parameters. The dataset was divided into two sub-datasets, namely, non-rainy and rainy. Principal component analysis (PCA) and factor analysis (FA) techniques were applied to evaluate seasonal correlations of water quality parameters and extract the most significant parameters influencing stream water quality. The first five principal components identified by PCA techniques explained greater than 80% of the total variance for both datasets. PCA and FA results indicated that total nitrogen, nitrate nitrogen, total phosphorus, and dissolved inorganic phosphorus were the most significant parameters under the non-rainy condition. This indicates that organic and inorganic pollutants loads in the streams can be related to discharges from point sources (domestic discharges) and non-point sources (agriculture, forest) of pollution. During the rainy period, turbidity, suspended solids, nitrate nitrogen, and dissolved inorganic phosphorus were identified as the most significant parameters. Physical parameters, suspended solids, and turbidity, are related to soil erosion and runoff from the basin. Organic and inorganic pollutants during the rainy period can be linked to decayed matters, manure, and inorganic fertilizers used in farming. Thus, the results of this study suggest that principal component analysis techniques are useful for analysis and interpretation of data and identification of pollution factors, which are valuable for understanding seasonal variations in water quality for effective management.

Key words: factor analysis, Lake Doam watershed, principal component analysis

INTRODUCTION

Lake Doam, located in Daegwallyeong, Gangwon-do was formed as part of a constructed dam in the upper reaches of River Song at the highest elevation of 712 m in South Korea in 1990. The dam has generated hydroelectricity since August of the same year. Water withdrawn from the dam was transported to the hydroelectric power plant through a tunnel, which subsequently reduced water discharge in the downstream and degraded the water quality. Therefore, the electric power generation was halted in 2001 (Cho 2001). The watershed includes residential areas, forests, mid-upland agricultural lands, pastures, golf links, and recreational resorts. Pollutant discharges from point sources (domestic wastewater and wastewater treatment plants) and non-point sources (agricultural, urban runoff, and forest) have been a big concern for Lake Doam. Long-term trends of summer season water quality

by Kwak et al. (2013) showed that Lake Doam was a meso-eutrophic lake where the average values of chlorophyll-a (Chl-a), total nitrogen (TN), and total phosphorus (TP) were 11.8 mg m^{-3}, 3.13 mg L^{-1}, and 0.077 mg L^{-1}, respectively.

Non-point source pollution caused by stormwater runoff is difficult to manage and high values of suspended solids (SS), total phosphorus (TP), and chemical oxygen demand (COD) were observed from 2009 to 2011 (Heo 2012). Agricultural activities in alpine areas with steep slope and soil conditioning practices are the major causes of soil erosion and sediment yield in the watershed. And the use of chemical fertilizers and livestock manure is a great concern of non-point source pollutant discharge, as it deteriorates the ecology and environment of water bodies (Shin 2004, Heo et al. 2006, Heo et al. 2007). Park et al. (2012) demonstrated that contamination in the water of Lake Doam and watershed streams was serious. Soil conservation to reduce SS and appropriate management are needed in the Lake Doam watershed to improve the water quality during wet weather conditions (Choi et al. 2012). Water quality degradation of water bodies due to runoff from highland agricultural areas is a major problem in South Korea (Heo et al. 1998, Park et al. 2005, Kim et al. 2007, Jung et al. 2009).

Water body contamination in rivers and lakes is one of the major global environmental problems. Pollution control is necessary to maintain good quality of water bodies. Therefore, a monitoring program is needed to provide a representative and reliable estimation of the quality of surface water bodies. And data analysis is required to interpret and understand the nature of pollution (Dixon and Chiswell 1996). The surface water sampling over time at various locations within the watershed for multiple physical, chemical, and biological water quality parameters is needed for a comprehensive water quality investigation. Multivariate statistical methods are applied to reduce the large and complex dataset matrices to better understand the water quality and ecology of the studied systems (Olsen et al. 2012). Principal component analysis (PCA) and factor analysis (FA) techniques are frequently applied to identify important components or factors of water variables. Water quality variables accounting for higher coefficients or loadings in the same principal component (PC) are interrelated and interpretation of PCs will help to understand temporal and spatial variations. This process of grouping variables and finding major parameters enables the identification of the sources of contamination (Thurston and Spengler 1985, Cao et al. 1999, Alberto et al. 2001, Simeonov et al. 2003, Ouyang et al. 2006, Shrestha and
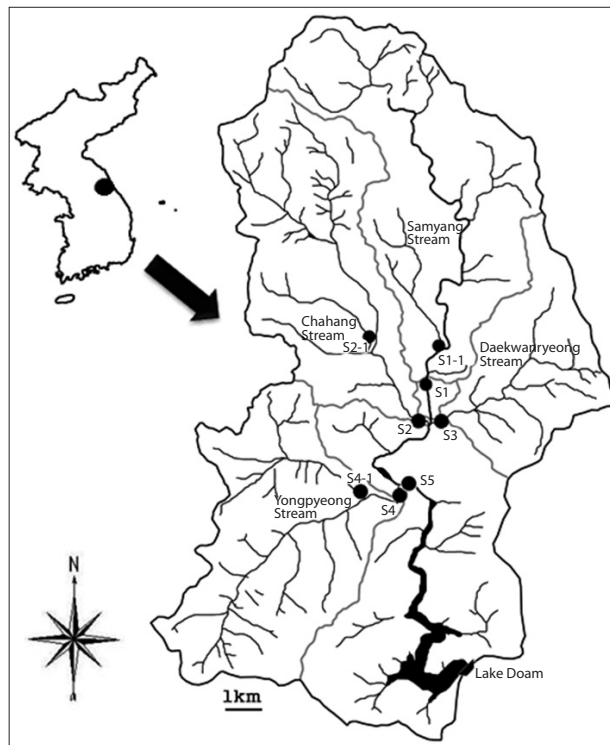

Fig. 1. Map of study area and monitoring sites in Lake Doam watershed. The study area's location on the Korean peninsula is shown as a dot on the map in the upper left corner. The geographic coordinates of the study area are lat 38°35'-38°47'N and long 128°37'-128°47'E.

Table 1. Major rainfall events at Lake Doam watershed from 2010-2013

Year	No. of rainfall events	Rainfall period	Total rainfall (mm)
2010	5	May 18-19	17.5
		May 22-24	81.0
		August 10-11	19.0
		September 2	18.5
		September 10-11	68.5
2011	4	June 23-24	133.0
		June 29-30	69.0
		July 7-8	61.0
		July 27-28	58.3
2012	8	April 2-4	64.7
		June 30-July 1	17.0
		July 19-20	35.5
		August 12-13	67.0
		August 15-16	18.5
		August 28-29	36.0
		August 30-31	92.5
		September 17-18	99.5
2013	5	May 27-28	26.2
		July 11-12	13.5
		July 15-16	122.0
		July 23-24	16.0
		September 29-30	25.0

Kazama 2007). PCA and FA methods have been applied to analyze and assess the water quality of several water bodies around the world over the last decade: River Elbe, Germany (Petersen et al. 2001); Gomti River, India (Singh et al. 2005); Jajrood River, Iran (Razmkhah et al. 2010); Qiantan River, China (Huang et al. 2010); Kinta River, Malaysia (Gazzaz et al. 2012); Tigris River basin, Turkey (Varol et al. 2012); and the Alqueva reservoir, Portugal (Palma et al. 2014).

This study aims to assess temporal and spatial variations of water quality parameters monitored from 2010 to 2013 in the Lake Doam watershed. Applying the multivariate statistical methods, PCA and FA, will make it easier to understand major water quality components or factors affecting water bodies in the watershed. The results of this work will help water managers to understand major pollutants and sources of pollution in the Lake Doam watershed.

MATERIALS AND METHODS

Study areas

River Song, the main stream flowing into Lake Doam, is 52.06 km in length. Tributaries of River Song are the Samyang Stream (15.67 km), Chahang Stream (10.45 km), Daegwallyeong Stream (17.34 km), and Yongpyeong Stream (8.60 km). The basin area (119 km^2) consists of sub-watershed areas: Samyang (42 km^2), Chahang (29.6 km^2), Daegwallyeong (16.8 km^2), and Yongpyeong (30.6 km^2). Lake Doam watershed is made up of forested areas (56%), agricultural areas (36%), and other areas (8%; residential infrastructures, roads, and streams). These eight sites (S1, S1-1, S2, S2-1, S3, S4, S4-1 and S5) were chosen

for the monitoring program (Fig. 1). According to the Wonju Regional Environmental Office's 2011 report, the total population in the area was 5,541 and the cattle population was 3,700 as of 2009. Alpensia and Yongpyeong ski resorts, which are located in this watershed, also attract large tourist populations.

Sampling and measurements

Sampling was conducted in two different weather conditions: such as rainy and non-rainy seasons. During the non-rainy season, water samples were collected once a month and tested, whereas during the rainy season, water samples were collected at the intervals of 1-2 hours in the rain. Overall, there were 22 notable rainfall events: 5, 4, 8, and 5 times in 2010, 2011, 2012 and 2013, respectively (Table 1).

The following water quality parameters were chosen for this study: water temperature (WT), electric conductivity (EC), pH, turbidity (Tur), suspended solids (SS), chemical oxygen demand (COD), total phosphorus (TP), dissolved inorganic phosphorus (DIP), total nitrogen (TN), nitrate nitrogen (NO_3-N), and ammonia nitrogen (NH_3-N) (Table 2). WT, EC, and pH were monitored at the sites using a multi-probe (YSI 556 MPS; YSI Incorporated, Yellow Springs, OH, USA). Water samples were collected in 2 L low density polyethylene (LDPE) bottles from each site and tested at the laboratory according to the standard method (American Public Health Association 1998) and the standard water pollution process test method (Institute of Environmental Research 1999). BRAN-LUEBBE Auto-Analyzer 3 (Bran Luebbe, Chicago, IL, USA) and UV-Vis spectrophotometer (Perkin-Elmer, Waltham, MA, USA) were used to test nitrogen and phosphorus components, respectively.

Table 2. Water quality parameters, units and analytical methods/instruments

Measured parameter	Symbol	Unit	Measuring instrument/method
Water temperature	WT	^{o}C	Multiprobe (YSI556MPS)
Electric conductivity	EC	μS/cm	Multiprobe (YSI556MPS)
pH	pH	pH unit	Multiprobe (YSI556MPS)
Turbidity	Tur	NTU	Turbidimeter (HACH 2100N)
Suspended solids	SS	mg/l	Glass fiber apparatus
Chemical oxygen demand	COD	mg/l	Water bath, KMNO4 acidic method
Total phosphorus	T-P	mg/l	Persulfate digestion, cadmium reduction, Ascorbic acid reduction
Dissolved inorganic phosphorus	DIP	mg/l	Ascorbic acid reduction
Total nitrogen	TN	mg/l	Alkaline persulfate digestion, Cadmium reduction
Nitrate nitrogen	NO_3-N	mg/l	Cadmium reduction
Ammonia nitrogen	NH_3-N	mg/l	Indophenol method

Statistical analysis

The large dataset was divided into two pools for the present study, rainy and non-rainy datasets, from all eight stations monitored during the four-year period of 2010 to 2013. The number of total observations in the non-rainy dataset was 360 (Table 3, sum of observations of 8 sites), while the number of total observations in rainy dataset was 1392 (Table 4, sum of observations of 8 sites). The multivariate statistical techniques, PCA and FA, were applied to each dataset to identify the most valuable components and factors to assess the water quality of the watershed. The PRINCOMP and FACTOR procedures were performed on Statistical Analysis System (SAS) software,

ver. 9.3 (SAS Institute Inc., Cary, NC, USA). The exploratory data analysis methods, PCA and FA, use the correlation structure among multiple parameters and produce a new set of variables that contain information on the original dataset. The PCA gives the eigenvalues of the correlation matrix, their proportions, and principal components as many as the parameters used in the dataset and corresponding eigenvectors of each component. The PC with the highest eigenvalue is known to be the most significant, and eigenvalues of 1.0 or greater are considered significant. The principal component can be expressed as:

$$z_{ij} = a_{i1}\chi_{1j} + a_{i2}\chi_{2j} + a_{i3}\chi_{3j} + \cdots + a_{im}\chi_{mj},$$

where z is the component score; a, the component load-

Table 3. Descriptive statistics of non-rainy dataset of Lake Doam watershed from 2010-2013

Site		WT	EC	pH	Tur	SS	COD	TP	DIP	TN	NO$_3$-N	NH$_3$-N
S1	Mean	15.1	83	7.3	15.6	13.8	3.8	0.081	0.020	2.678	2.243	0.075
	SD	6.7	36	0.9	23.2	18.9	2.3	0.204	0.040	1.038	0.972	0.070
	Max	27.1	206	8.6	86.6	82.0	11.0	1.342	0.266	6.494	5.099	0.241
	Min	1.4	24	4.3	1.4	1.8	1.2	0.003	0.002	1.294	0.003	0.002
S1-1	Mean	15.5	143	7.4	8.2	9.7	3.8	0.129	0.026	4.862	4.111	0.110
	SD	6.2	132	1.0	8.6	13.6	1.8	0.459	0.037	1.507	1.716	0.270
	Max	28.1	890	8.9	35.6	82.6	9.0	3.103	0.242	11.698	10.280	1.646
	Min	2.5	29	4.2	1.6	1.4	1.5	0.007	0.001	2.055	0.008	0.001
S2	Mean	15.8	205	7.3	12.6	13.6	4.0	0.180	0.048	5.883	5.161	0.092
	SD	6.2	69	0.8	14.5	18.4	2.0	0.619	0.046	1.291	1.428	0.113
	Max	27.2	331	8.6	56.2	79.4	9.6	4.210	0.308	8.656	7.660	0.498
	Min	2.9	50	4.6	1.7	1.5	1.6	0.018	0.011	3.059	0.025	0.001
S2-1	Mean	16.5	248	7.2	15.1	16.9	4.3	0.291	0.250	10.112	8.112	0.076
	SD	6.0	130	1.1	18.2	22.0	1.8	1.235	1.235	2.376	2.601	0.050
	Max	28.8	651	8.9	65.4	85.9	8.6	8.373	8.344	15.143	12.357	0.185
	Min	3.5	55	4.3	1.2	1.3	1.9	0.026	0.003	3.442	0.029	0.001
S3	Mean	14.8	184	7.3	8.5	9.4	3.5	0.169	0.039	4.969	4.355	0.082
	SD	5.2	120	1.0	9.7	11.7	2.3	0.640	0.048	0.882	1.093	0.051
	Max	25.0	840	8.7	41.6	47.5	13.6	4.342	0.337	6.634	5.835	0.235
	Min	3.9	49	4.1	1.3	1.2	1.3	0.023	0.007	3.232	0.033	0.003
S4	Mean	16.3	196	7.5	20.0	17.5	5.0	0.263	0.084	4.167	3.564	0.091
	SD	6.3	70	0.9	21.4	17.5	3.7	0.765	0.075	1.113	1.155	0.088
	Max	27.5	404	8.9	90.0	73.7	25.0	5.224	0.435	6.478	5.511	0.440
	Min	4.1	41	4.3	1.9	2.7	1.6	0.029	0.003	2.195	0.016	0.001
S4-1	Mean	16.3	241	7.4	26.7	25.0	6.1	0.227	0.065	5.579	4.802	0.087
	SD	6.4	97	0.9	26.1	22.6	3.5	0.452	0.055	1.352	1.388	0.116
	Max	26.3	577	9.0	95.0	86.5	25.8	3.047	0.363	9.584	8.245	0.679
	Min	3.7	50	4.5	2.6	2.1	2.3	0.039	0.006	2.881	0.027	0.001
S5	Mean	15.6	164	7.5	11.9	11.5	3.8	0.160	0.044	4.221	3.683	0.070
	SD	6.4	55	1.0	11.5	12.7	1.7	0.481	0.042	0.794	1.015	0.056
	Max	27.2	294	9.6	54.1	57.9	9.2	3.261	0.291	6.101	5.493	0.201
	Min	2.8	39	4.2	1.9	1.8	1.6	0.026	0.014	2.389	0.022	0.001
	N/site	45	45	45	45	45	45	45	45	45	45	36

*SD, standard deviation; Max, maximum value; Min, minimum value; N, number of observations; WT, water temperature; EC, electric conductivity; Tur, turbidity; SS, suspended solids; COD, chemical oxygen demand; TP, total phosphorus; DIP, dissolved inorganic phosphorus; TN, total nitrogen; NO$_3$-N, nitrate nitrogen; NH$_3$-H, ammonia nitrogen.

ing; χ, the meaused value of variable; i, the component number, j, the sample number; and m, the total number of variables.

Varifactors (VFs) are produced by further rotation of axis of new variables defined by the PCA. Factor analysis procedures follow the PCA for reduction of dataset to obtain only a few factors/components that have eigenvalues greater than 1.0. And it further reduces the contribution of less significant variables. The factors are valuable to get a good description of the entire dataset variability (Ouyang et al. 2006, Olsen et al., 2012). The expression of FA can be:

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \cdots + a_{fm}\chi_{mi} + e_{fi}$$

where z is the measured variable; a, the factor loading; f, the factor score; e, the residual term for errors; j, the factor number; i, the sample number; and m, the total number of factors.

Table 4. Descriptive statistics of rainy dataset of Lake Doam watershed from 2010-2013

Site		WT	EC	pH	Tur	SS	COD	TP	DIP	TN	NO$_3$-N	NH$_3$-N
S1	Mean	16.0	69	7.3	442.5	517.2	14.8	0.361	0.243	2.321	1.568	0.119
	SD	3.9	40	0.6	836.6	899.7	19.0	0.669	0.655	1.657	0.678	0.161
	Max	23.5	460	9.1	5000.0	5860.0	137.6	5.521	4.312	11.658	4.622	1.520
	Min	0.7	20	3.9	10.1	14.2	2.1	0.007	0.001	0.008	0.497	0.001
	N	193	194	193	178	179	195	195	194	195	195	157
S1-1	Mean	15.7	95	7.4	504.1	581.2	18.8	0.346	0.269	3.719	2.699	0.130
	SD	3.7	84	0.6	923.9	1082.6	28.2	0.560	0.737	2.150	1.202	0.198
	Max	22.1	889	9.8	5000.0	7323.0	178.8	4.723	4.491	12.750	9.395	1.697
	Min	1.6	29	5.1	8.9	11.3	1.4	0.014	0.001	0.006	0.544	0.001
	N	193	195	192	176	180	195	195	195	195	195	156
S2	Mean	16.1	139	7.3	546.0	643.0	18.9	0.533	0.617	4.826	3.862	0.185
	SD	3.8	77	0.6	837.8	974.0	23.9	0.809	1.790	2.546	1.654	0.202
	Max	23.2	796	9.3	5000.0	5704.0	154.8	8.258	9.984	18.752	11.818	1.319
	Min	1.1	16	4.9	10.3	14.8	2.8	0.037	0.012	0.014	1.123	0.001
	N	193	195	193	181	184	195	195	195	195	195	157
S2-1	Mean	16.1	162	7.4	623.7	687.3	18.9	0.617	0.761	7.898	6.692	0.244
	SD	3.8	62	0.6	932.2	984.6	23.0	0.892	2.154	3.637	2.776	0.348
	Max	23.1	541	10.2	5000.0	6466.0	154.0	8.997	11.436	21.976	16.342	2.506
	Min	1.4	50	5.2	11.4	14.2	1.7	0.028	0.004	0.026	0.897	0.001
	N	193	195	193	181	182	195	195	195	195	195	156
S3	Mean	15.5	128	7.2	483.5	460.2	14.6	0.400	0.388	3.741	2.948	0.171
	SD	3.6	67	0.5	882.2	788.3	19.1	0.635	1.026	1.839	1.064	0.162
	Max	21.5	810	8.9	5000.0	5219.0	171.8	5.891	7.916	11.754	6.849	1.049
	Min	1.4	30	4.8	8.4	10.4	1.5	0.026	0.019	0.008	1.110	0.002
	N	193	195	193	174	174	195	195	195	195	195	157
S4	Mean	16.7	157	7.4	467.5	468.8	12.6	0.448	0.365	3.750	2.918	0.160
	SD	4.0	86	0.6	759.9	717.7	12.2	0.628	0.868	1.737	1.084	0.132
	Max	24.5	851	8.8	4023.0	4492.0	75.2	7.343	4.848	10.158	7.041	0.942
	Min	2.1	66	4.8	13.6	16.7	2.2	0.044	0.004	0.014	0.732	0.010
	N	193	195	193	184	184	193	195	195	195	195	156
S4-1	Mean	17.4	179	7.3	593.9	711.5	16.6	0.559	0.462	4.580	3.390	0.184
	SD	3.9	56	0.5	840.8	1192.6	19.2	0.767	1.270	2.401	1.429	0.208
	Max	23.8	329	8.8	5000.0	8185.0	131.2	8.100	9.870	16.328	9.926	1.670
	Min	1.9	16	4.7	12.0	14.7	2.2	0.049	0.001	0.028	0.498	0.006
	N	193	195	193	182	190	195	195	195	195	195	157
S5	Mean	16.5	118	7.3	477.3	505.9	16.1	0.419	0.377	3.404	2.701	0.140
	SD	3.9	46	0.5	863.1	823.9	21.9	0.643	1.004	1.757	0.949	0.125
	Max	23.2	291	8.4	5000.0	5317.0	174.4	6.652	5.870	12.722	6.472	0.998
	Min	1.2	28	4.9	9.4	10.9	1.3	0.028	0.019	0.020	0.904	0.010
	N	193	194	193	176	179	195	195	195	195	195	157

*SD, standard deviation; Max, maximum value; Min, minimum value; N, number of observations; WT, water temperature; EC, electric conductivity; Tur, turbidity; SS, suspended solids; COD, chemical oxygen demand; TP, total phosphorus; DIP, dissolved inorganic phosphorus; TN, total nitrogen; NO$_3$-N, nitrate nitrogen; NH$_3$-H, ammonia nitrogen.

RESULTS AND DISCUSSION

The non-rainy and rainy datasets in Table 3 and 4 provide descriptive statistics (mean, standard deviation, etc.) of the monitored parameters for all eight sites during the four-year period of 2010 to 2013. These two tables show a similar pattern for all the sites: mean values of parameters, EC, TN, and NO_3-N, were higher during the non-rainy condition, and Tur, SS, COD, TP, DIP, and NH_3-N were higher during the rainy condition. Lower mean values of EC, TN, and NO_3-N during the rainy condition could be the result of dilution during high flow. The highest mean turbidity was observed at site S2-1, with a value of 624, followed by sites S4-1 of 594 and S2 of 546. Site S4-1 had the highest mean value of SS, with a value of 712, and sites S2-1 and S2 also recorded high mean values of SS, with mean values of 687 and 643, respectively. Sites S2 and S2-1 observed the highest COD. Phosphorus components TP and DIP were recorded high at site S2-1 with the mean values of 0.617 and 0.761, respectively. Nitrogen components TN, NO_3-N, and NH_3-N were also higher at site S2-1 and the mean values were 10.11, 8.11, and 0.244, respectively. These parameters were also recorded high in two other sites, S2 and S4-1. The two sites, S2-1 and S2, which

recorded high mean values of water quality parameters, are located in the Chahang stream and S4-1 is located in the Yongpyeong stream. According to Heo (2012), nearly 97% of the total cultivated land of the Lake Doam watershed lies in the sub-watersheds of Chahang and Yongpyeong streams. The Chahang stream sub-watershed and Yongpyeong stream sub-watershed are made up of about 62% and 52% of cultivated land, respectively. Soil erosion and agricultural runoff could be the main sources of the high values of water quality parameters in these two stream sites. The rainy data observations in this study were mostly monitored from June to September, which is the monsoon season in Korea. The monsoon brings large and intense rainfall in this area. Soil erosion from the watershed is the major cause of water turbidity water and high concentrations of SS and pollutants (Heo et al. 1998). Upper stream sites at the tributaries of the Song Stream receive direct discharge from natural and anthropogenic point and non-point sources of pollution, and obvious influences were shown in the results.

Data in Table 5 demonstrate the correlation matrices for both rainy and non-rainy datasets. Overall, correlations between parameters were very low in both datasets. However, in the non-rainy dataset, SS and Tur, and TN and

Table 5. Correlation matrix of the 11 water quality parameters monitored in Lake Doam watershed

	WT	EC	pH	Tur	SS	COD	TP	DIP	TN	NO₃-N
Non-rainy										
WT	1									
EC	-0.0751	1								
pH	**0.4452**	-0.2695	1							
Tur	-0.0987	0.0935	-0.1320	1						
SS	-0.1419	0.1320	-0.1665	**0.8307**	1					
COD	0.1558	-0.0104	0.1344	0.3649	0.3024	1				
TP	-0.1478	-0.0406	0.0762	0.0272	0.0220	0.0236	1			
DIP	-0.0060	0.0436	-0.0393	0.0250	0.0399	0.0549	-0.0048	1		
TN	-0.0164	**0.4356**	-0.1923	0.0681	0.1366	0.0260	-0.0434	0.1778	1	
NO₃-N	-0.0133	0.4177	-0.2023	0.0857	0.1692	0.0279	-0.2596	0.1744	**0.9107**	1
Rainy										
WT	1									
EC	0.0647	1								
pH	0.0225	0.0102	1							
Tur	0.0983	-0.0497	-0.2199	1						
SS	-0.0337	-0.0052	-0.2945	**0.8063**	1					
COD	0.0337	-0.0378	-0.2522	0.6670	**0.7314**	1				
TP	0.0569	-0.0704	0.0152	0.5632	0.3926	0.1651	1			
DIP	-0.0147	-0.0834	-0.0683	0.0867	0.0668	-0.0611	0.0936	1		
TN	0.0784	0.3755	-0.0603	0.2395	0.2111	0.2265	0.1929	-0.3436	1	
NO₃-N	0.0385	**0.4054**	-0.0257	0.0314	0.0380	0.0025	-0.0031	0.1827	**0.7314**	1

*WT, water temperature; EC, electric conductivity; Tur, turbidity; SS, suspended solids; COD, chemical oxygen demand; TP, total phosphorus; DIP, dissolved inorganic phosphorus; TN, total nitrogen; NO_3-N, nitrate nitrogen.
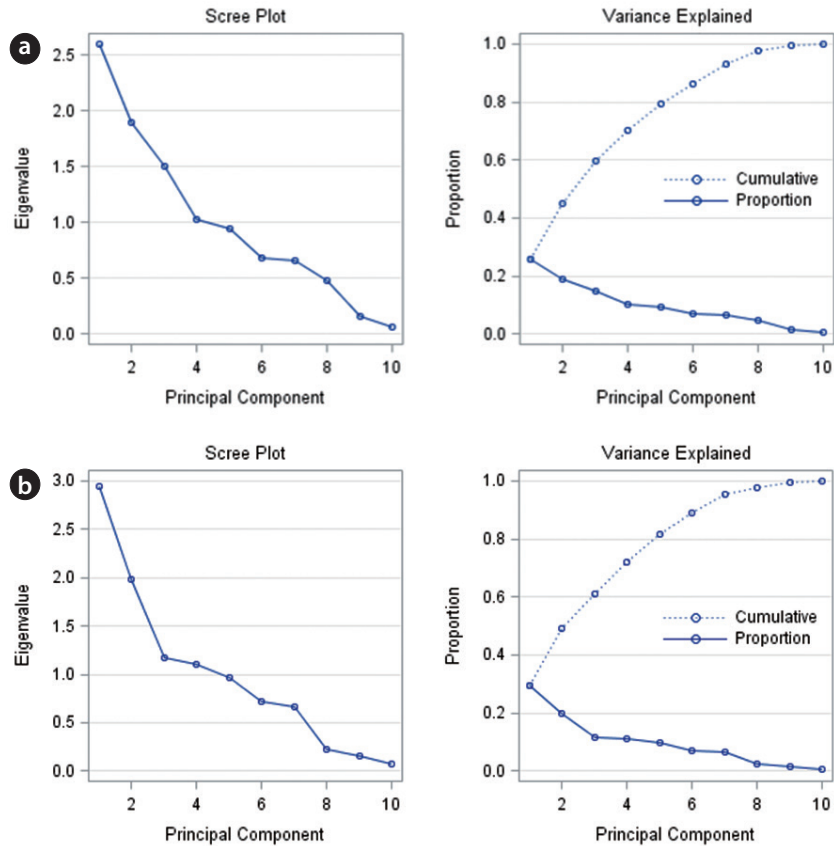
Fig. 2. Eigenvalues of principal components. (a) Non-rainy, (b) Rainy datasets.

NO$_3$-N, showed strong significant correlations of 0.83 and 0.91, respectively. Among the parameters of the rainy dataset, SS had highly significant positive correlations with Tur and COD, which were 0.81 and 0.73, respectively. Significant positive correlations were also recorded between parameters, Tur and COD, and TN and NO$_3$-N, of 0.67 and 0.73, respectively. The high correlation between COD and SS is a very interesting case. Discharge from organic pollutants in the stream from natural and anthropogenic sources is the cause of high COD (Simeonov et al. 2003, Singh et al. 2005). Stormwater runoff from residential areas, cattle farms, forest, and the manure used on agricultural lands could be the main reason behind the positive correlation between SS and COD.

The results demonstrated by the PCA procedure were presented in Fig. 2 and 3. Eigenvalues, proportions, and variance explained by each principal component and the loadings on the first five components were presented in these two figures. The eigenvalues accounted for the first five PCs of non-rainy data, which were 2.60, 1.90, 1.50, 1.03, and 0.94. And the eigenvalues for the first five PCs of rainy data were 2.95, 1.99, 1.17, 1.11, and 0.97. The total

variances explained by the first five PCs of non-rainy and rainy datasets were 80% and 81%, respectively. Therefore, the first five PCs for both datasets that had eigenvalues greater than 1.0 or near to unity are only considered for further analysis.

Component loadings of the first four PCs with eigenvalues of greater than 1.0 were retained and presented in Fig. 3 for both monitored conditions. The first principal component of the non-rainy dataset (Fig. 3a) explained 26% of the total variance and was largely and positively contributed by TN and NO$_3$-N. Moderate contributions were from EC, SS, and Tur. This factor represents high nutrient pollutants of organic and in-organic nature from point and non-point sources of pollution. And the moderate loadings on EC, SS, and Tur are from the physical source of variability. Strong and positive loadings on Tur and SS were observed in PC2 along with positively loaded variations of COD and TP. Therefore, PC2 represents erosion related to physical and organic sources of pollutants. PC3 had strong and positive loadings on WT, pH, and COD due to physical and organic sources of pollution. And PC4 loadings were mostly dominated by TP, DIP, and
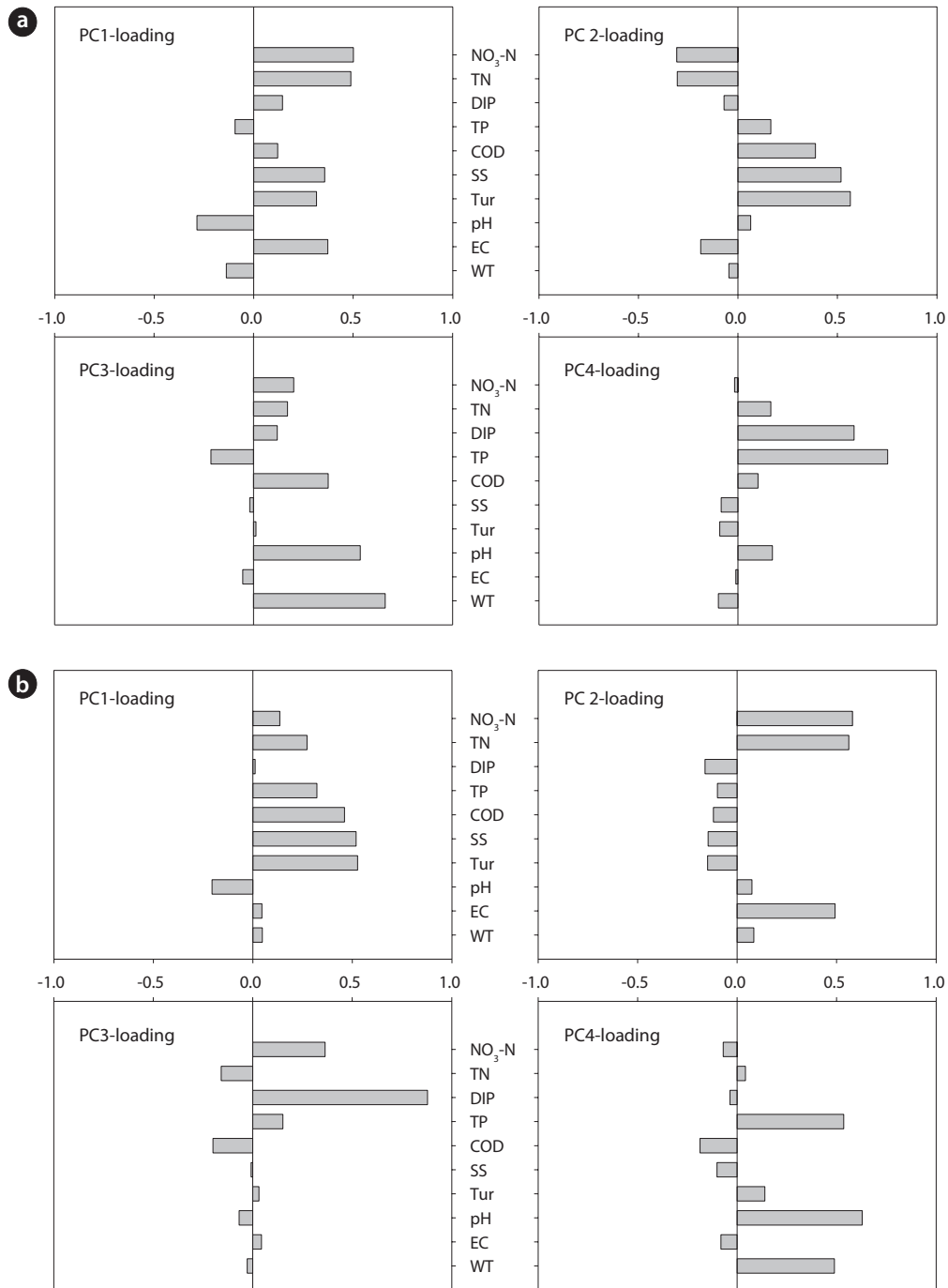
Fig. 3. Component loadings for four principal components of datasets. (a) Non-rainy, (b) Rainy. PC, principal component; NO₃-N, nitrate nitrogen; TN, total nitrogen; DIP, dissolved inorganic phosphorus; TP, total phosphorus; COD, chemical oxygen demand; SS, suspended solids; Tur, turbidity; EC, electric conductivity; WT, water temperature.

TN, which can be interpreted as representing influences from point sources and non-point sources. The sources of organic and in-organic pollutants in the stream are discharges from domestic wastewater, wastewater treatment plants, cattle farms, and agricultural lands. Park et al. (2005) reported that TP and TN were also recorded high in the Seomjin River, which receives inflows from paddy fields and dry fields. The variances explained by PC2, PC3, and PC4 of non-rainy data were 19%, 15%, and 10%, respectively.

Unlike the case of the PC1 of non-rainy data, the PC1 of rainy data (Fig. 3b) was largely contributed by Tur, SS, and COD. Small positive contributions in PC1 were noted from nutrient constituents TN and TP, and negative from pH. The high loading on suspended solids indicates the soil erosion from upland areas during the rainy season and the positive correlations with chemical oxygen demand indicates the presence of partially decayed organic matter in washout from forests and agricultural areas (Shrestha and Kazama 2007). This component reveals that WT, EC, and DIP were less important. The PC1 of rainy dataset had a slightly higher variance than the PC1 of non-rainy data, which was 30%. The PC2 of rainy dataset explained 19% of the total variance and was largely and positively contributed by EC (mineral component) and nitrogen components, TN and NO_3-N. The PC3 showed strong positive loadings on in-organic related parameters, DIP and NO_3-N. The PC4 was loaded on organic source parameters, pH and TP, and the physical parameter WT.

Component loadings for the first five principal components of both datasets showed that all PCs were negatively and positively influenced by most of the variables, which makes it quite difficult to interpret which parameters were more important than others in affecting water quality. Therefore, factor analysis is needed to better understand the major parameters influencing water quality during non-rainy and rainy conditions. The FA process assigns higher correlation coefficients for the significant parameters and lower correlation coefficients for the less significant parameters.

Varimax rotated factor patterns for the first five PCs, which accounted for more than 80% for both datasets, were presented in Table 6 and Table 7. In this study, any water quality parameter with an absolute correlation coefficient value of >90% was considered to be an important parameter and greater than 80% was considered to be a moderately important parameter. The absolute value of the correlation coefficient >95% was considered by Ouyang et al. (2006) as important, but Shrestha and Kazama (2007) considered >75% as important. In the non-rainy dataset, TN and NO_3-N were the most important parameters of Factor 1 and their correlation coefficients were 0.93 and 0.91, respectively. TP and DIP had correlation coefficients of 0.98 and 0.97 in Factor 4 and Factor 5, respectively. Tur and SS in Factor 2 and WT in Factor 3 also had significant correlation coefficients (Table 6). Factor 1 of the rainy dataset showed significant correlation coefficients of Tur, SS, and COD, which were 0.93, 0.91, and 0.80, respectively. The most important parameters in Factor 2 of the rainy data were NO_3-N (0.91) and TN (0.84) with positive correlations. The other highly significant coefficients were shown in Factor 3 by DIP and in Factor 5 by WT (Table 7).

The variations of most important parameters identified by a principal component method were presented in Fig. 4 and Fig. 5. Spatial distributions presented in box and whisker plot trends give clear information that the sites with high variations of the major parameters were S2-1,

Table 6. Varimax rotated factor pattern of non-rainy dataset

Parameters	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
EC	**0.68753**	0.05500	-0.18051	0.06173	-0.15317
TN	**0.93387**	0.03402	0.00276	0.00797	0.15186
NO_3-N	**0.90584**	0.06205	-0.00970	-0.21871	0.15050
Tur	0.03136	**0.92541**	-0.12366	-0.00756	-0.02082
SS	0.10910	**0.89708**	-0.17899	-0.01095	-0.00956
COD	0.00866	**0.61029**	0.40146	0.06651	0.09694
WT	0.02865	-0.04673	**0.82698**	-0.19739	-0.03401
pH	-0.21490	-0.07431	**0.79832**	0.15455	-0.00533
TP	-0.06999	0.02174	-0.03173	**0.97892**	0.00348
DIP	0.08546	0.02459	-0.02693	0.00350	**0.97409**
Variance explained by each factor					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
2.237513	2.050395	1.564135	1.077446	1.029140

*EC, electric conductivity; TN, total nitrogen; NO_3-N, nitrate nitrogen; Tur, turbidity; SS, suspended solids; COD, chemical oxygen demand; WT, water temperature; TP, total phosphorus; DIP, dissolved inorganic phosphorus.

Table 7. Varimax rotated factor pattern of rainy dataset

Parameters	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Tur	**0.92906**	0.02318	0.07809	0.07660	0.08434
SS	**0.90807**	0.03549	0.03501	-0.12693	-0.06776
COD	**0.80140**	0.00981	-0.16785	-0.27601	0.01273
TP	**0.59972**	0.00804	0.17312	**0.59926**	0.04484
NO_3-N	0.00420	**0.90888**	0.24283	0.00501	-0.02904
EC	-0.10504	**0.68838**	-0.07000	-0.08317	0.07521
TN	0.27475	**0.84157**	-0.30985	0.09676	0.00951
DIP	0.02097	-0.04979	**0.97844**	-0.03855	-0.00644
pH	-0.31124	-0.02868	-0.11894	**0.77598**	0.00664
WT	0.02557	0.05330	-0.00652	0.01991	**0.99502**
Variance explained by each factor					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
2.874171	2.016270	1.196874	1.077605	1.010623

*Tur, turbidity; SS, suspended solids; COD, chemical oxygen demand; TP, total phosphorus; NO_3-N, nitrate nitrogen; EC, electric conductivity; TN, total nitrogen; DIP, dissolved inorganic phosphorus; WT, water temperature.
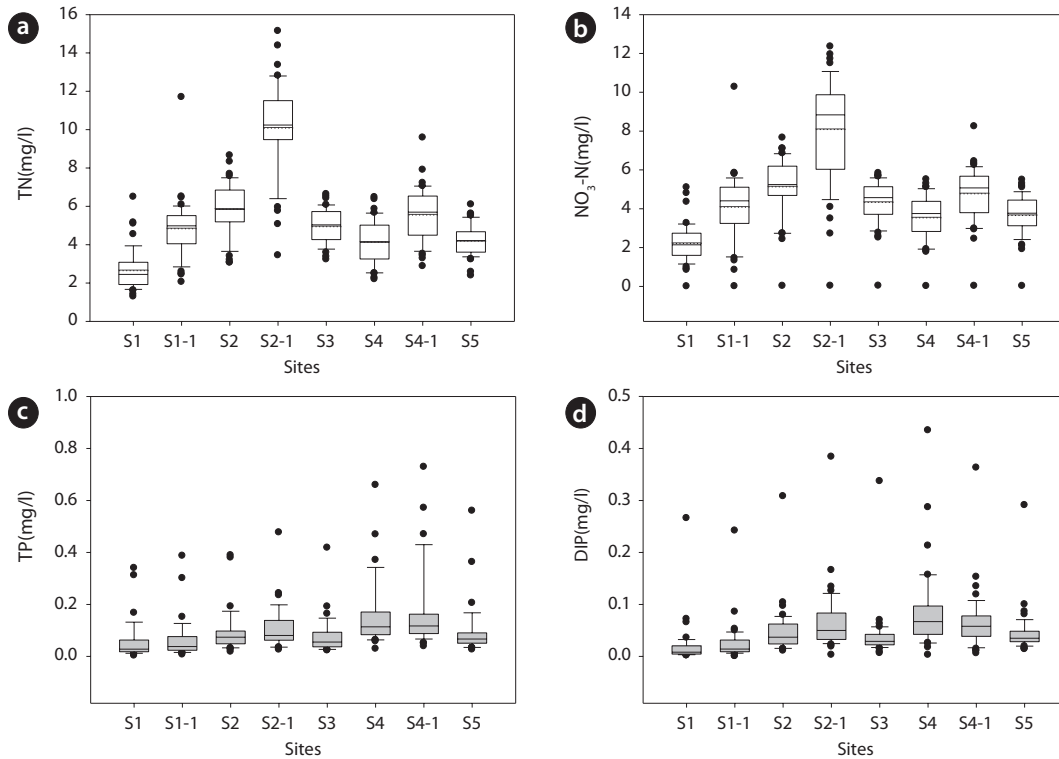
Fig. 4. Box plots of four important parameters (TN, NO₃-N, TP, DIP) identified by factor procedure in non-rainy conditions. (a) TN, (b) NO₃-N, (c) TP, (d) DIP. TN, total nitrogen; NO₃-N, nitrate nitrogen; TP, total phosphorus; DIP, dissolved inorganic phosphorus; S, site. Mean, dotted line; median, dark line; box, 1st and 3rd quartiles; whiskers, mean ± standard deviation; outliers, dark points.
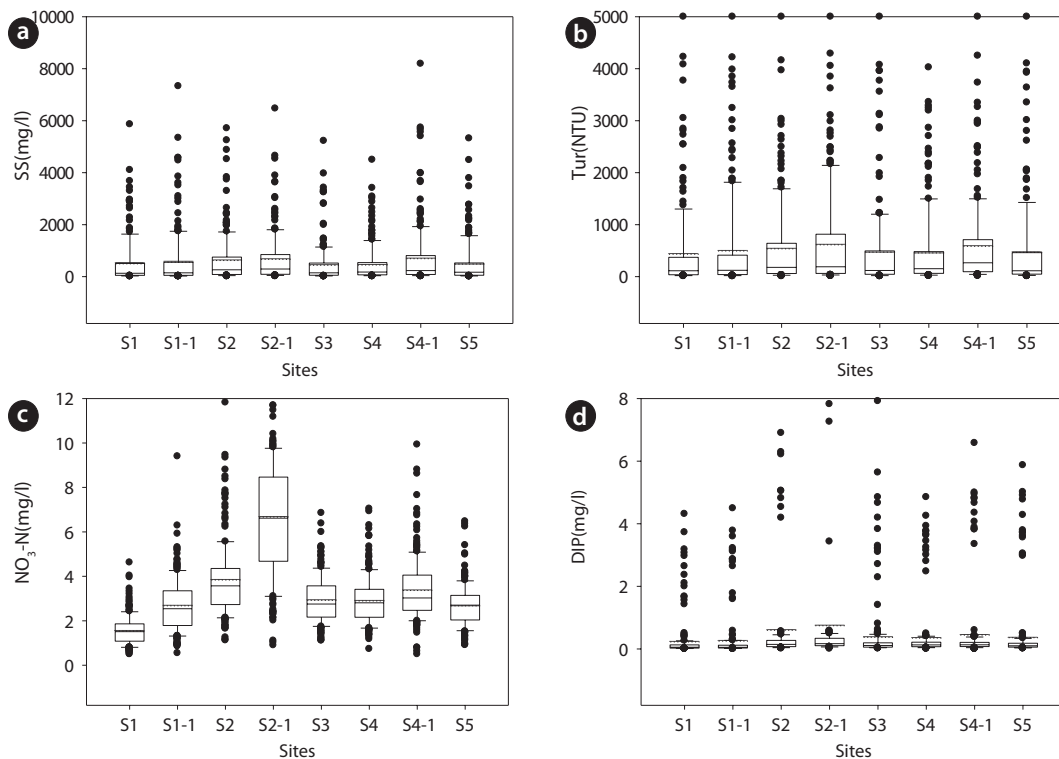

Fig. 5. Box plots of four important parameters (SS, Tur, NO₃-N, DIP) identified by factor procedure in rainy conditions. (a) SS, (b) Tur, (c) NO₃-N, (d) DIP. SS, suspended solids; Tur, turbidity; NO₃-N, nitrate nitrogen; DIP, dissolved inorganic phosphorus; S, site. Mean, dotted line; median, dark line; box, 1st and 3rd quartiles; whiskers, mean ± standard deviation; outliers, dark points.

S4-1, S2, and S4. These four sites are located in the upper reaches of Lake Doam watershed and receive discharge from sub-watersheds, which are mainly dominated by forest and agricultural areas. The main reason for high concentration of variables can be erosion and runoff related discharge of organic and inorganic pollutants in the streams. During the non-rainy conditions, nitrogen components (TN, NO_3-N) showed higher variations at sites S2-1, S4-1 and S2, and phosphorus components (TP, DIP) showed higher variations at sites S4-1 and S4 than the other sites. Nitrate nitrogen (NO_3-N) variations were high during rainy conditions at sites S2-1, S4-1, and S2, too. During rainy conditions, high variations of SS, Tur, and DIP were observed at sites S2, S2-1 and S4-1.

This study demonstrated that the water quality variations of Lake Doam watershed under non-rainy and rainy conditions were distinctly different. The highly significant parameters (total nitrogen and total phosphorus) identified in the non-rainy dataset were less significant in the rainy dataset. Turbidity, suspended solids, and chemical oxygen demand were highly significant in the rainy dataset but less significant in the non-rainy dataset. Therefore, water managers and administrators must consider non-rainy and rainy seasonal water quality variations while selecting parameters to reduce pollutant loads.

CONCLUSION

This study used principal component analysis techniques to analyze and interpret the water quality variations of streams in the Lake Doam watershed during non-rainy and rainy conditions. Principal component analysis demonstrated that the first five principal components represented 80% of the total variance explained in each dataset. Varifactors obtained from the factor analysis process indicate that the significant parameters influencing water quality in one condition were not entirely same in the other condition. As presented in the results, nitrate nitrogen, total nitrogen, total phosphorus, and dissolved organic phosphorus were the most important variables for non-rainy conditions, while suspended solids, turbidity, nitrate nitrogen, and dissolved inorganic phosphorus were the most important variables for rainy conditions. The pollutants loading into the streams during the non-rainy conditions can be related to point source pollution and non-point source pollution from the watershed. Soil erosion and decayed matter related pollutants were dominantly loaded during rainy conditions. The spatial distribution of the FA identified that the most important

parameters had higher concentrations at the sites of Chahang and Yongpyeong tributaries than the other sites. These sites receive discharges mainly from cultivated areas in the uplands. Thus, the principal analysis techniques can be helpful to analyze and interpret the water quality data and information can be used to assess the water quality of water bodies in the watershed.

LITERATURE CITED

Alberto WD, Del Pilar DM, Valeria AM, Fabiana PS, Cecilia HA, De Los Angeles BM. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba–Argentina). Water Res 35: 2881-2894.

American Public Health Association, American Water Works Association, Water Environment Federation. 1998. Standard methods for the examination of water and wastewater, 20th ed. Washington, D.C., American Public Health Association.

Cao Y, Williams DD, Williams NE. 1999. Data transformation and standardization in the multivariate analysis of river water quality. Ecol Appl 9: 669-677.

Cho JH. 2001. The impact of hydroelectric power plant discharge on the water quality of the Namde River. Environ Eng Res 6: 123-130.

Choi JW, Shin DS, Lim KJ, Lee SS, Kang MJ. 2012. Estimation suspended solids concentration of the Doam Reservoir under dry and wet weather conditions. Kor J Environ Agric 31: 113-121.

Dixon W, Chiswell B. 1996. Review of aquatic monitoring program design. Water Res 30: 1935-1948.

Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, Juahir H. 2012. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. Mar Poll Bull 64: 688-698.

Heo SG, Kim JY, Yoo DS, Kim KS, Ahn JH, Yoon JS, Lim KJ. 2007. Simulation of hydrological and sediment behaviors in the Doam-dam watershed considering soil properties of the soil reconditioned agricultural fields. J Korean Soc Agr Eng 49: 49-60.

Heo SG, Kim KS, Ahn JH, Yoon JS, Lim KJ, Choi JD, Shin YC, Lyou CW. 2006. Evaluation of water quality impacts of forest fragmentation at Doam-Dam watershed using GIS-based modeling system. J Kor Assoc Geo Info Stud 9: 81-94.

Heo WM, Kim BC, Kim YH, Choi KS. 1998. Storm runoff of phosphorus from nonpoint sources into Lake Soyang and transportation of turbid watermass within the lake.

Kor J Limnol 31: 1-8.

Heo WY. 2012. Characteristics of non-point pollution discharge on stormwater runoff from Lake Doam watershed. Kor J Ecol Environ 45: 62-71.

Huang F, Wang X, Lou L, Zhou Z, Wu J. 2010. Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. Water Res 44: 1562-1572.

Institute of Environmental Research. 1999. Standard method for examination of water and wastewater. Ministry of Environment, Sejong, Republic of Korea.

Jung SM, Jang CW, Kim JK, Kim BC. 2009. Characteristics of water quality by storm runoffs from intensive highland agriculture area in the upstream of Han River basin. J Kor Soc Water Qual 25: 102-111.

Kim JK, Shin MS, Jang CW, Jung SM, Kim BC. 2007. Comparison of TOC and DOC distribution and the oxidation efficiency of BOD and COD in several reservoirs and rivers in the Han River system. J Kor Soc Water Qual 23: 72-80.

Kwak SJ, Bhattrai BD, Lee CK, Heo WM. 2013. Long-term trends of summer season of water quality in Lake Doam. Kor J Ecol Environ 46: 128-134.

Olsen RL, Chappell RW, Loftis JC. 2012. Water quality sample collection, data treatment and results presentation for principal components analysis-literature review and Illinois River watershed case study. Water Res 46: 3110-3122.

Ouyang Y, Nkedi-Kizza P, Wu QT, Shinde D, Huang CH. 2006. Assessment of seasonal variations in surface water quality. Water Res 40: 3800-3810.

Palma P, Ledo L, Soares S, Barbosa IR, Alvarenga P. 2014. Spatial and temporal variability of the water and sediments quality in the Alqueva reservoir (Guadiana Basin; southern Portugal). Sci Tot Environ 470-471: 780–790.

Park KH, Kim BS, Yun HJ, Ryu KY, Yun JC, Choi JY, Kim KD, Jin YI. 2012. Characteristics of water quality at main streams and Lake Doam in Daegwallyeong area. Korean J Soil Sci Fert 45: 882-889.

Park SC, Oh CR, Jin YH, Kim DS. 2005. Study on runoff characteristics of non-point source in rural area of Seomjin watershed. J Environ Sci 14: 1057-1062.

Petersen W, Bertino L, Callies U, Zorita E. 2001. Process identification by principal component analysis of river water-quality data. Ecol Model 138: 193-213.

Razmkhah H, Abrishamchi A, Torkian A. 2010. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). J Env Manag 91: 852-860.

Shin YK. 2004. Comparison of water quality between forested and agricultural subcatchments in Daegwallyong area. J Korean Assoc Geogr Stud 39: 544-561.

Shrestha S, Kazama F. 2007. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Model Softw 22: 464-475.

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimtzi TH. 2003. Assessment of the surface water quality in Northern Greece. Water Res 37: 4119-4124.

Singh KP, Malik A, Sinha S. 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques-a case study. Analytica Chimica Acta 538: 355-374.

Thurston GD, Spengler JD. 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. Atmosph Environ 19: 9-25.

Varol M, Gökot B, Bekleyen A, Sen B. 2012. Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. Catena 92: 11-21.

Wonju Regional Environmental Office. 2011. Doam dam non-point pollution source management, area monitoring and evaluation program. Ministry of Environment, Sejong, Korea; 11-1480358-000051-01.